THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 61, No. 1, January 1982
Printed in U.S.A.

A Sequential Projection Algorithm for Special-Services Demand

By A. IONESCU-GRAFF

(Manuscript received December 31, 1980)

Performance analysis of the four time-independent regression models presently used by Bell operating companies to forecast special-services circuit requirements, and the characteristics of actual special-services demand history observed from three operating companies, indicate a need for a new method to forecast these difficult time series. A new special-services demand sequential projection algorithm (SSD-SPA) is developed based on a linear Kalman filter model. It includes methods to detect previous deterministic events, to accept and process exogenous information affecting the demand, and to recognize and adapt to a "no-growth" situation. Compared to the present algorithm, SSD-SPA generates significantly better forecasts: approximately 30 percent improvement in forecast accuracy and stability, 25 percent reduction in rms error, and 22 percent reduction in circuit misplacements.

I. INTRODUCTION AND SUMMARY

In recent years, the demand for special-services circuits has grown at more than twice the annual rate of the demand for message telephone service (9 percent versus 4 percent). This rapid growth, the development of new technologies, and the problems in the existing special-service provisioning process have led to a reexamination of the overall process of special-services planning and provisioning.

Key inputs to this process are special-services demand forecasts; they are required for the marketing, budgeting, and engineering functions. Presently, in most Bell operating companies (BOCS), the short-range forecast (1 to 5 years) of point-to-point demands for interoffice special-service circuits is provided by forecasting systems based on time-independent trending models or by applying a user specified growth factor to the most recent demand.

Previous studies of the special-service circuits life-time distribution

found no single, common distribution that reasonably fit the observed data. The time series consist mainly of small integers with demand levels ranging from very volatile to perfectly constant, and displaying numerous "jumps," probably the effects of deterministic events. These data characteristics explain the inadequacy of the present time-independent (unweighted) regression models used to fit the past data: linear, exponential, and first- and second-order autoregressive.

Consequently, a new algorithm—the special services demand sequential projection algorithm (SSD-SPA)—is proposed, based on a dynamic time-series model with deterministic event input, the Kalman filtering technique for state-vector estimation and prediction, and an additional procedure to process outliers. The attributes and specific parameters of this model are derived from the demand history for special services from three BOCs.

Section II gives background information on the study. It describes the data available for analyses, summarizes the main characteristics of the demand time series to be forecasted, and presents the measures to be used in the empirical investigation of the algorithms' performances. A brief overview of the existing forecasting models, and results of the forecasting algorithm performance analysis follow. A list of the desirable features of a new special-services demand projection algorithm are derived from these results and the characteristics of the actual demand history mentioned in Section 2.1.

In Section III, a linear Kalman filter model is formulated, and the choice of specific parameters is studied. Implementation considerations include initialization, outlier detection, deterministic event (level or growth) detection and processing, as well as filter gain selection. Special-services demand sequential projection algorithm forecasts are then tested and compared to the present forecasting algorithm. Results include the comparative forecast qualities for the case of small integer projection and an estimated economic impact of the new algorithm. Finally, conclusions and recommendations are summarized in Section IV.

II. BACKGROUND

Evaluation and comparison of various demand projection algorithms require a description of the characteristics of the time series to be projected, so that the appropriateness of the model can be determined; also, a definition of the performance statistics used for algorithm comparison is needed, so that the best feasible model can be selected.

2.1 Special-services demand

2.1.1 Data for analyses

The term *special services* refers to all Bell System services other than ordinary message telephone service. Examples of special services

are foreign exchange, tie lines, off-premise extensions, and private lines. The classification of special-service circuits varies from one BOC to another, and for a single BOC over time. For example, one BOC recognizes about 500 different circuit types, while another recognizes only 150.

Two types of special-services history files were available from three major BOCs to support our studies: detailed-demand history files (DDHF) and grouped-demand history files (GDHF). The maximum number of available months varied among the three BOCs: 60 for Company A, 67 for Company B, and 71 for Company C. The maximum could not exceed 71 months since this is the maximum that can be stored and processed by the present forecasting system.

The DDHF contains *individual* records of the number of special-service circuits of a given BOC class of service between a pair of central offices (COs). The large number of possible point-to-point individual special-service type circuit records on the DDHF (for example, 500 types for each pair of COs times all possible combinations of CO pairs) and the small size of these groups (more than 90 percent have only one circuit) makes any attempt to forecast each time series impractical. Consequently, in the design of the present forecasting system (the special-services forecasting system, or SSFS), the decision was made to group the individual records before projection.

This grouping of DDHF records, according to a user specified grouping strategy, results in a GDHF. The resultant grouped special-service time series are the basic input to the forecasting routines and represent the numbers of circuits of one or more types between a given pair of offices.

For the special-services demand analysis, both types of files were used. For the present forecasting algorithm performance study, only the GDHFs from Companies B and C could be used since only they had the format required by the input routine. These two files were also used for the SSD-SPA performance tests.

Both tapes were created using grouping strategies specified by the facility and equipment planners: 14 grouping types in Company B and 19 types in Company C. The file from Company B covers the time period between January, 1973 and July, 1978 and contains 20,036 such grouped records. The file from Company C extends over the period January, 1973 to November, 1978 and contains 41,073 records.

2.1.2 Demand characteristics

The special-services demand analysis identified the following significant characteristics:

(*i*) Very skewed circuit group size distribution, regardless of the grouping strategy. More than 80 percent of the point-to-point groups

consist of less than 10 circuits. Fig. 1 plots the maximum number of circuits in service over the history for each circuit group against the frequency of that particular size. The histograms for the three BOCs are remarkably similar, even though the grouping strategies used were different. The skewness of the size distribution would be accentuated if we had plotted the group sizes at a given point in time, instead of the maximum size over the whole history. In the same time, the long tail of the distribution shows that, although most of the point-to-points are very small in size, the majority of the special-service circuits are placed in a few very large groups. For example, in Company C only 6.5 percent of the groups consist of more than 50 circuits, but these groups are extremely large and account for more than 75 percent of all special circuits in service.

(*ii*) No seasonal pattern.

(*iii*) High volatility of the time series, even at high levels of aggregation.

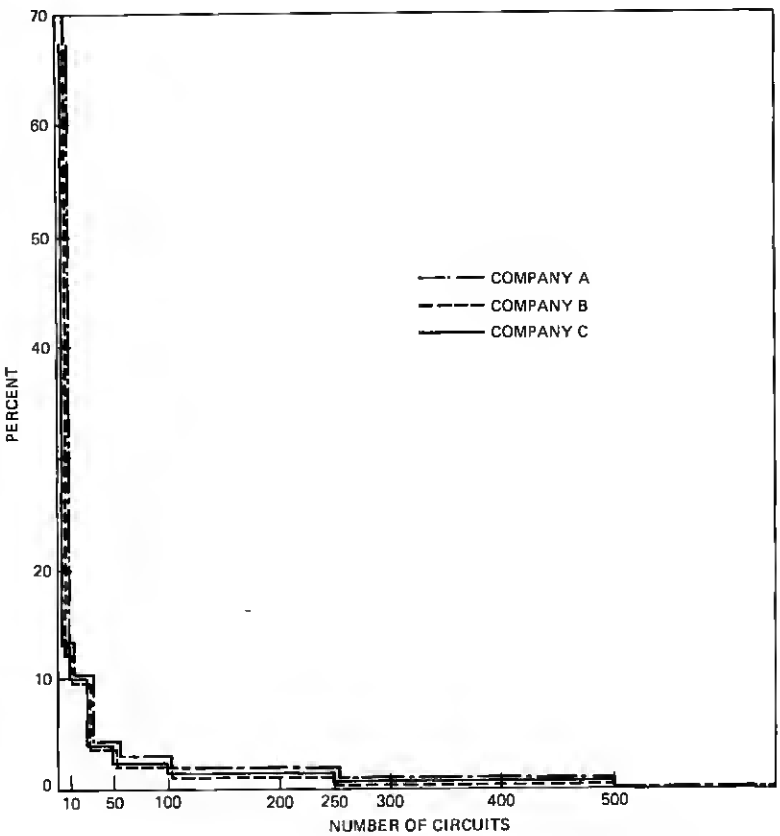

Fig. 1—Maximum demand per circuit forecasting group.

(*iv*) Jumps in the demand level. This is a frequent phenomenon; circuit groups remain at one size for an extended period of time, then jump to another value, and remain at this second level for some time. This stepwise change in the demand level is probably caused by deterministic events, such as large customers moving in or out, routing changes, tariff changes, or market stimulation.

(*v*) Constant level circuit groups. Approximately 40 percent of each company's grouped point-to-point records showed no change in the demand level over the period of time that data were available.

(*vi*) Vanishing circuit groups. About 30 percent of the groups have all their circuits eventually disconnected (i.e., the demand for these groups goes to zero) with practically no regeneration during the period.

2.2 Algorithm performance measures

Previous studies[1,2] indicated that good performance measures for algorithm comparisons are accuracy (average forecast error), rms error (square root of the mean squared error), and stability (the variability of consecutive views of the same future period). For the present analysis, a fourth forecast attribute, misplacement, is defined (total positive or negative forecast error).

To quantitatively measure and compare the performance of both SSFS and SSD-SPA forecasting procedures, we used both algorithms to generate demand forecasts for each circuit group (from the same data base), and then compared the results using relative forecast error statistics.

Let:

$y_{n+k} \equiv$ the recorded number of special-service circuits at time $n + k$

$\hat{x}_{n+k,n} \equiv$ the forecasted demand at time $n + k$, given data through time n; i.e., a k-period forecast.

Then the relative accuracy, $\hat{A}_{n+k,n}$, of the k-period forecast from period n is defined to be

$$\hat{A}_{n+k,n} = \left(\frac{\hat{x}_{n+k,n} - y_{n+k}}{y_{n+k}}\right). \tag{1}$$

The relative rms error, $\hat{R}_{n+k,n}$, is defined to be the square root of

$$\hat{R}^2_{n+k,n} = \left[\frac{\hat{x}_{n+k,n} - y_{n+k}}{y_{n+k}}\right]^2. \tag{2}$$

The relative stability, $\hat{S}_{n+k,n}$, of consecutive forecasts from periods $n - 1$ and n for a fixed target date $n + k$ is defined by:

$$\hat{S}^2_{n+k,n} = \left[\frac{\hat{x}_{n+k,n} - \hat{x}_{n+k,n-1}}{y_{n+k}}\right]^2. \tag{3}$$

Treatment of $y_{n+k} = 0$ is discussed later. Accuracy and stability are actually measuring inaccuracy and instability. Consequently, a decrease in either measure is equivalent to an improvement. All three performance measurements are empirical estimates of the normalized statistics (accuracy, rms error, and stability) as described in Ref. 2. Relative statistics, as opposed to absolute statistics, were used so that a small absolute error on a large group would not obscure large absolute errors on many other small groups. Network estimates of accuracy, stability, and rms error are produced by averaging individual estimates over all groups.

We define total error ($\hat{\text{TE}}$) to be

$$\hat{\text{TE}} = \frac{\text{Total forecast} - \text{total demand}}{\text{Total demand}}.$$

Note that $\hat{\text{TE}}$ can be almost zero as a result of error cancellations; therefore, misplacement is a better measure of the total number of circuits erroneously forecasted. Misplacements translate directly into inefficient capital expenditures.

The positive misplacement, $\hat{M}^{+}_{n+k,n}$, of the total number of circuits forecasted from period n for the target period $n + k$ is defined by:

$$\hat{M}^{+}_{n+k,n} = \left(\frac{\sum_{i=1}^{N} d_i}{\sum_{i=1}^{N} y^{(i)}_{n+k}} \right), \tag{4}$$

where $i = 1, 2, \cdots, N$ is the index over all circuit groups in the network and

$$\begin{aligned} d_i &= \hat{x}^{(i)}_{n+k,n} - y^{(i)}_{n+k} \quad \text{if} \quad \hat{x}^{(i)}_{n+k,n} \geq y^{(i)}_{n+k} \\ &= 0 \qquad\qquad\qquad\qquad \text{otherwise.} \end{aligned}$$

Similarly, the negative misplacement, $\hat{M}^{-}_{n+k,n}$, of the total number of circuits forecasted from period n for the target period $n + k$ is defined by:

$$\hat{M}^{-}_{n+k,n} = \left(\frac{\sum_{i=1}^{N} d_i}{\sum_{i=1}^{N} y^{(i)}_{n+k}} \right),$$

where $i = 1, 2, \cdots, N$ is the index over all circuit groups in the network and

$$\begin{aligned} d_i &= y^{(i)}_{n+k} - \hat{x}^{(i)}_{n+k,n} \quad \text{if} \quad \hat{x}^{(i)}_{n+k,n} < y^{(i)}_{n+k} \\ &= 0 \qquad\qquad\qquad\qquad \text{otherwise.} \end{aligned}$$

We call $M^+_{n+k,n}$ a measure of total network overprovisioning and the negative misplacement, $M^-_{n+k,n}$, a measure of total network underprovisioning. Positive misplacement may translate into underutilization, while negative misplacement may translate into orders lost or held, or misroutings.

Note that TE and misplacements are related as

$$\text{TE} = M^+_{n+k,n} - M^-_{n+k,n} .$$

2.3 Prasent forecasting algorithm

This section presents a brief overview of the projection algorithm presently used in SSFS, the performance testing procedure, and its results.

2.3.1 Overview

The present forecasting algorithm produces point-to-point demand forecasts of interoffice special-services circuits for the current year and for each of the next 5 years.

The forecast is generated in two major steps—the preliminary forecast and the final forecast. The preliminary forecast employs one of four statistical models or user-stated growth factors to predict future circuit requirements. The four regression models are linear, exponential, and first- and second-order autoregressive. They are used only when the group has sufficient demand history; at least 12 months of history are always required, and the default value is 24 months. Before forecasting, the available history is smoothed using a 3-month moving average.

The parameters for each model are determined by minimizing an unweighted sum of squared errors over the smoothed data. The model with the smallest sum of squared errors or, equivalently, the model with the highest R^2 statistic (the coefficient of determination of "goodness of fit"), is selected. However, the exponential model is rejected if any of the history is zero or if it would lead to a prediction of explosive growth, and the autoregressive models are rejected, unless the demand time series is sufficiently stationary.

Finally, if the model chosen was linear or exponential and the current demand has shifted significantly from the historical growth trend, then the forecast is also shifted to coincide with the current demand. A significant shift is defined relative to the estimated standard error of the unsmoothed demand history (excluding the current demand) from the trend line. Since at each forecast view all history is reprocessed to recalculate the regression parameters, treatment of such discontinuities may be inconsistent from one forecast view to another. The sensitivity of this test may be adjusted by the user; the

default value is two standard deviations. No adjustment of this kind is considered for the autoregressive models.

When the forecast groups do not have the required number of months of history, forecasts are produced applying default growth factors to the forecast group's current demand.

The forecaster reviews tbe preliminary forecast and makes manual adjustments when appropriate. An example is when advance knowledge is available on new businesses moving into an area or new services are being offered.

The following section describes the results of our study to quantify the present forecasting algorithm performance. This analysis only covers the preliminary forecast. The impact of manual adjustments was not studied, since no records were available. The main deficiencies of the existing forecasting technique are summarized and explained in view of the demand time series characteristics.

2.3.2 Performance analysis

The algorithm performance is specified in terms of statistical attributes (accuracy, rms error, stability, misplacements); the analysis sought to verify if there is indeed a benefit in having four different models to choose from, to identify the main forecasting problem, and, based on the demand time series characteristics, to derive requirements for a new forecasting algorithm.

A modified version of SSFS was used to produce up to three consecutive forecasts for each point-to-point demand, depending on the length of eacb demand history available. To ensure compatibility with other planning tools, SSFS is required to produce quarterly average forecasts of the future demand for special services. The data files available extend up to 71 months, and since SSFS requires at least 12 months of history for the forecast initialization, the longest forecast that can be produced and checked against actuals is 18 to 19 quarters, i.e., about 4 years. For simplicity, instead of estimating 18 to 19 values of $\hat{A}$, $\hat{R}$, $\hat{S}$, $\widehat{\text{TE}}$, and $\hat{M}$, we only looked at one quarter in each year (the same quarter each year, right justified by the last quarter of available data). Consequently, for those records with at least 60 months of history, three forecasts were provided, as shown in Fig. 2 (1 year initialization plus 4-year-span forecast, then 2 years initialization and 3-year-span forecast, and 3 years initialization and 2-year-span forecast). Only two forecasts were produced for records with 48 to 59 months of history (3- and 2-year-span forecasts), and only one forecast for records with 36 to 47 months of history.

Since each forecast is made after at least 12 months of data are processed, only a steady-state analysis is necessary. Thus, the subscripts n for $\hat{A}_{n+k,n}$, $\hat{R}_{n+k,n}$, and $\hat{S}_{n+k,n}$ are dropped. For each circuit

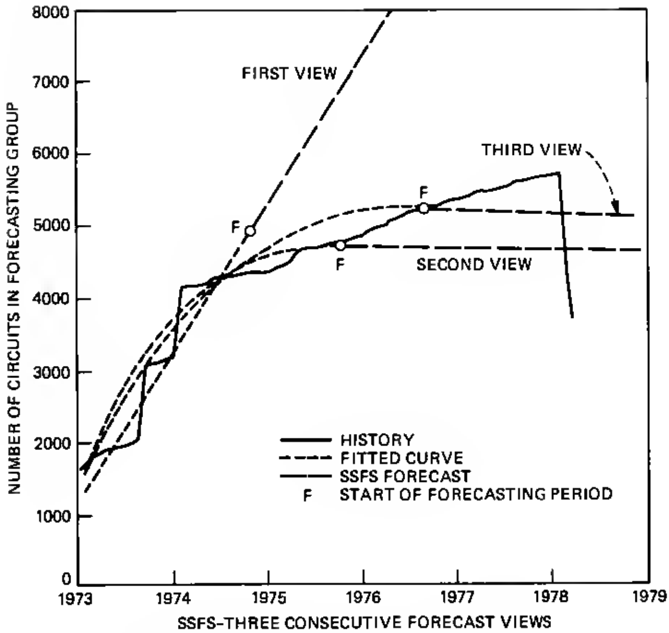

Fig. 2—Algorithms performance analysis test plan.

group, accuracy, rms error, and stability are estimated using all forecasts produced. For example, the accuracy of a 3-year-ahead forecast for a circuit group with 60 months of history available is estimated by:

$$\hat{A}_3 \triangleq \frac{1}{2}\left(\frac{\hat{x}_{77,74} - y_{77}}{y_{77}} + \frac{\hat{x}_{78,75} - y_{78}}{y_{78}}\right),$$

where, if q is the last quarter of available data in the last year of history, then

$\hat{x}_{i,j}$ = forecast of the average demand in quarter q; year i, made from quarter q; year j.

$\hat{y}_i$ = average measurements of the demand in quarter q; year i.

Or, stability of 3- versus 4-year-ahead forecasts for the same group is estimated by:

$$\hat{S}_3 \triangleq \left|\frac{\hat{x}_{78,74} - \hat{x}_{78,75}}{y_{78}}\right|.$$

For groups where $\hat{y}_{n+k} = 0$, a normalization factor of 1 is used. This may bias the statistics (to look worse than they actually are), but since the objective is to *compare* the performance of different algorithms, and they all use this rule, we may expect this normalization to affect

them equally. In fact, the performance results measured using this normalization were similar to those obtained with nonvanishing groups only. Three consecutive forecasts were produced for about 56 percent of the groups (those groups with 60 to 67 months of history*), two forecasts for about 10 percent (48 to 59 months of history), and only one forecast for 9 percent (36 to 47 months of history). The remaining 25 percent of the groups were not used, since their recorded histories were shorter than 36 months.

2.3.2.1 Statistical performance. The results showed that the demand forecasts are often inaccurate and unstable. The numerical results can be deduced from the values presented in Section 3.3 on the SSD-SPA performance, and relative improvement versus the present SSFS.

The accuracy histogram showed about 40 percent of the groups had 1-year forecasts with no error. This was to be expected since about 40 percent of the point-to-point groups have constant demand. Additionally, the existing forecasting algorithms more often overforecasted than underforecasted.

The significant instability observed for consecutive forecasts was due not only to the intrinsic volatility of the demand time series, but also to the change in forecasting models used each year.

Small total errors resulted from cancellations of up to 50 percent misplaced circuits (large total misplacement). It was interesting to observe that although accuracy was always positive, many times (especially for Company C) the total error was negative. This means that even if on the average most of the circuit groups are overforecasted, some of the very large point-to-point groups are underforecasted so that the total forecast over the whole company is less than the realized demand.

2.3.2.2 Correlation between forecasting model fit (R^2) and projection error (accuracy and rms). As previously described, the existing algorithm chooses from the four regression models the one with the highest coefficient of determination (R^2, or "goodness of fit"). The intuitive reason for this is that the curve that best fits the past data should extrapolate most accurately into the future. If indeed, there is a benefit in having four different models from which to choose, then we would expect to find some correlation between how well the chosen models fit the data (R^2) and the forecast quality. Subsequent testing, designed to consider all combinations of models and forecasting spans, showed that the choice of four projection models appeared unjustified since the correlation between the goodness of the model fit to the history

* The number of months of history refers to how long ago the first circuits were installed on that group, not to the actual length of time the demand was nonzero. About 30 percent of the groups with more than 36 months of recorded history vanished during that period (demand had zero value eventually).

data (R^2) and the forecast errors (accuracy or rms) was statistically insignificant. In other words, even perfect knowledge of the past does not necessarily imply good knowledge of the future.

2.3.2.3 Outlier detection procedure. Many of the demand time series display a stepwise, highly volatile growth pattern, with the jumps probably generated by deterministic events. The existing shift option reacts to a significant difference between the actual demand and the forecasted value only if it happens in the last month of history. Any other jumps in previous months are treated as normal trend. Moreover, the error monitoring capability, which can detect large forecast deviations from the actual demand, is exterior to the main forecasting process. Consequently, the next projection cannot be improved based on the detected past errors. Figs. 3a and 3b give examples when the wrong model or parameters were selected for projection because of improper treatment of past special events.

Another deficiency is the rather slow response to changes in trend; the equal weight assigned to each history point prevents the system from adjusting itself quickly to recent changes.

2.3.2.4 Manual adjustments. The present forecasting algorithm cannot accept and process exogenous information. The forecaster has to review manually the forecasts and supply any modifications based on up-to-date knowledge. For example, about 70 percent of the forecasts in Company C are manually adjusted.

2.3.2.5 History requirements. The special-services forecasting system needs at least 12 months (usually 24 months) of history to produce a forecast based on one of the four statistical models. If less data are available, growth factors are used (default or manually input values);

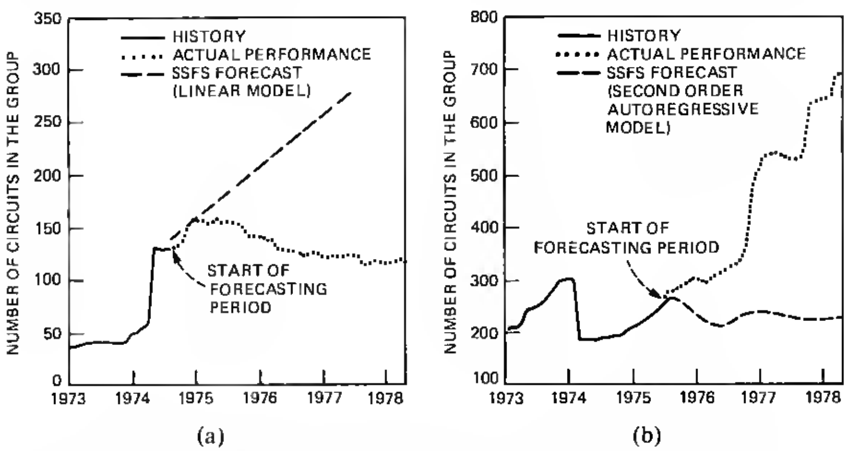

Fig. 3—Circuit groups with deterministic "jumps." SSFS forecasting problem: special event treated as normal growth. (a) Example 1; (b) Example 2.

15.2 percent of the Company B data base and 17 percent of the Company C data base consist of circuit group records with less tban 24 months of history.

2.3.3 New algorithm requirements

The demand time series cbaracteristics and the results from the present algorithm's performance suggest some desirable properties of a new algorithm:

(*i*) Unequal weighting of data. Weigh tbe most recent data more heavily to allow tbe forecasting algorithm to adapt to dynamic changes in the demand pattern.

(*ii*) Acceptance of exogenous information. Point-to-point demand levels are significantly affected by special events, such as large customers moving in or out, tariff changes, or by market stimulation. Many of these events are known in advance and their impact on the individual time series can be estimated. The forecasting system should accept those estimates and use tbem in projecting future demand levels.

(*iii*) Shorter initialization period. The special-services segment of the total Bell System network is constantly changing and growing. New technologies, services, and rates are changing the customer demand patterns. Many special-service circuit groups are eventually disconnected (on an individual basis, 50 percent of the special-services circuits bad a lifetime of less than 36 months; and 30 percent of the total number of groups "died" over a 5-year period); other groups come into service. Thus, a forecasting system must produce accurate forecasts based on small amounts of historical data, e.g., less than 12 months.

(*iv*) Recognition of past deterministic events (step changes and constant levels). The system should be able to recognize and react to "significant" changes in the demand level. Significant has to be defined as a function of the observed demand time series characteristics.

(*v*) Forecast of small integers. About 80 percent of the special-services circuit groups have less than 10 circuits in service. Whatever the model selected for projection, it should produce stable and accurate forecasts of integers from 1 to 10.

(*vi*) Computational efficiency. Users find it useful to run SSFS on a monthly basis.

III. SPECIAL-SERVICES DEMAND SEQUENTIAL PROJECTION ALGORITHM

A linear dynamic system with linear growth and deterministic input is shown to be a reasonable and robust approximation for the special-services demand time series, and a simple (two-dimensional) linear

Kalman filter is selected as a method to estimate the state variables of this system. Filter parameter selection is examined and procedures to detect and respond appropriately to outliers are added to capture the stepwise growth pattern of the demand time series. Using data from Companies B and C, we test the performance of this new algorithm and compare it with the present algorithm.

3.1 Linear Kalman filter model

3.1.1 Model formulation

In a linear dynamic model, as discussed in Reference 2, the behavior of the discrete time series is determined by an s-dimensional state-vector process $\{\mathbf{X}_n\}$. The following two equations describe the time evolution of the process $\{\mathbf{X}_n\}$ and the relation between $\mathbf{X}_n$ and the corresponding observation $\mathbf{y}_i$:

$$\text{System equation: } \mathbf{X}_{n+1} = \phi_n \mathbf{X}_n + \mathbf{U}_{n+1} + \omega_{n+1} \tag{5}$$

$$\text{Observation equation: } \quad \mathbf{y}_n = H_n \mathbf{X}_n + \nu_n \,, \tag{6}$$

where ϕ_n is an $s \times s$ state transition matrix, ω_n is an s-dimensional modeling error vector, $\mathbf{U}_n$ is an s-dimensional deterministic input, H_n is a $d \times s$ observation matrix, and ν_n is a s-dimensional measurement noise vector. Furthermore,

$$E(\nu_n) = E(\omega_n) = 0$$

$$E(\omega_n \omega_j^T) \equiv \begin{cases} 0 & \text{if } \quad n \neq j \\ Q_n & \text{if } \quad n = j \end{cases} \qquad Q_n \text{ an } s \times s \text{ matrix}$$

$$E(\nu_n \nu_j^T) \equiv \begin{cases} 0 & \text{if } \quad n \neq j \\ R_n & \text{if } \quad n = j \end{cases} \qquad R_n \text{ an } d \times d \text{ matrix}$$

$$E(\nu_n \omega_j^T) \equiv 0 \qquad \text{for all } n, j.$$

The $s \times s$ matrix, Q_n, is known as the modeling error covariance matrix and R_n is the measurment error covariance matrix.

In our demand analysis, it was demonstrated that no single common demand pattern exists for special services, but that for the majority of groups a linear model fit the historical data best. Furthermore, earlier work using Kalman filters for forecasting message trunk group loads,[1,3] showed that for short-term forecasting applications a linear Kalman filter model performs well even for nonlinear processes such as an exponential.* Consequently, we chose to develop a linear model that accounts for the special-demand characteristics discussed earlier.

---

* Reference 1, for example, analyzed the performance of different Kalman filter models (linear, log-linear, etc.) to forecast busy season trunk group loads. It showed that, given measurement and modeling errors and errors in the initial state estimates, the linear Kalman filter would produce short-term (1 to 5 years) forecasts as good as any other model with respect to accuracy, rms, and stability measures.

Given the univariate measurement time series (i.e., $d = 1$) with linear growth, the special-services demand model can be represented by a two-dimensional linear model with the following parameters:

$$\phi_n \equiv \phi = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}; \qquad H_n = (1, 0); \tag{7}$$

$$\mathbf{X}_n = \begin{pmatrix} x_n^1 \\ x_n^2 \end{pmatrix}, \tag{8}$$

where x_n^1 represents the demand level at time n, and x_n^2 the incremental growth.

3.1.2 Kalman filter (filtering and prediction)

The Kalman filter is a recursive method that produces a minimum variance unbiased estimate of the state vector $\{\mathbf{X}_n\}$ of a dynamic linear system from noisy observations $\mathbf{y}_1, \cdots, \mathbf{y}_n$ and uses these estimates to predict future state values.

Let $\hat{\mathbf{X}}_{n,n-i}$ be the estimate of the state vector $\mathbf{X}_n$ based on information available through time $n - i$. Let

$$P_n = E[(\mathbf{X}_n - \hat{\mathbf{X}}_{n,n-1})(\mathbf{X}_n - \hat{\mathbf{X}}_{n,n-1})^T]$$

be the one-step prediction error covariance matrix and

$$S_n = E[(\mathbf{X}_n - \hat{\mathbf{X}}_{n,n})(\mathbf{X}_n - \hat{\mathbf{X}}_{n,n})^T]$$

be the estimation error covariance matrix. Then, given a prior estimate of the system state $\hat{\mathbf{X}}_{n,n-1}$, the filtering problem is to find an updated estimated $\hat{\mathbf{X}}_{n,n}$, based on the measurement $\mathbf{y}_n$.

The unbiased estimate is given by the linear recursive form

$$\hat{\mathbf{X}}_{n,n} = \hat{\mathbf{X}}_{n,n-1} + K_n(\mathbf{y}_n - H_n\hat{\mathbf{X}}_{n,n-1}), \tag{9}$$

where K_n is a time-varying weighting matrix known as the Kalman filter gain matrix. The optimal* K_n is given by

$$K_n = P_nH_n^T(H_nP_nH_n^T + R_n)^{-1}. \tag{10}$$

The error covariance matrices are found to be:

$$S_n = (I - K_nH_n)P_n, \tag{11}$$

where I is the $s \times s$ identity matrix, and

$$P_{n+1} = \phi_nS_n\phi_n^T + Q_n. \tag{12}$$

---

* The criterion of optimality is to minimize the mean square estimation error. When ω_n and ν_n are normally distributed, then the same result is obtained by a Bayesian method or the method of maximum likelihood.

The estimates of the future state vectors are obtained by extrapolation using eq. (5)

$$\hat{\mathbf{X}}_{n+k,n} = \phi_{n+k-1}\hat{\mathbf{X}}_{n+k-1,n} + \mathbf{U}_{n+k}$$
$$= \left(\prod_{l=0}^{k-1} \phi_{n+l}\right) \hat{\mathbf{X}}_{n,n} + \sum_{l=1}^{k-1} \mathbf{U}_{n+l} \left(\prod_{m=l}^{k-1} \phi_{n+m}\right) + \mathbf{U}_{n+k} . \qquad (13)$$

It should be mentioned that if ω_n and ν_n are Gaussian, the Kalman filter estimate is at least as good as any other estimate (either linear or nonlinear). If the noise terms cannot be assumed normal, then the Kalman filter yields the optimal linear unbiased minimum variance estimate, but there may be a nonlinear estimate that is superior in mean square error.[4,5]

To implement the above described algorithm, we note that:

(*i*) An initial state estimate and error covariance are necessary to start the recursion. This problem will be considered in Section 3.2.1.

(*ii*) Since the estimation error covariance matrix S_n and prediction error covariance matrix P_n do not rely on observed data, for given sequences $\{Q_n\}$, $\{R_n\}$, and initial P_{n_0},* the gain sequence $\{K_n\}$ can be precalculated. Specification of Q_n and R_n will be discussed in Section 3.2.2. The choice of the gain sequence will be examined in Section 3.2.3.

(*iii*) It is not necessary to store the history $\{\mathbf{y}_1, \cdots, \mathbf{y}_n\}$ since all relevant information concerning the series is included in the state vector estimate $\hat{\mathbf{X}}_{n,n}$.

(*iv*) The algorithm assumes the knowledge of the future deterministic events $\{\mathbf{U}_n\}$. If estimates of the impact of these events are not available (user input) or are in error, the system needs a recovery procedure. However, when a significant change in the demand is observed, the algorithm has to differentiate between outliers (because of measurement errors, or demand volatility) and deterministic events. The problems of outlier detection and response to special events are considered in Sections 3.2.4 and 3.2.5.

3.2 Implementation considerations and parameter selection

In most applications, the exact statistical structure of the individual time series is unknown. Consequently, implementation of the Kalman filter model requires selection of estimated values for the algorithm parameters, usually through experimentation. Three methods to obtain initial estimates for $\hat{\mathbf{X}}_{n_0,n_0-1}$ and $\hat{P}_{n_0}$ will be analyzed next. Then, the specification of R and Q and the choice of gains and outlier thresholds will be considered.

* Time n_0 is the assumed "present time" for filter initialization, given the available data history $\{y_1, \cdots y_{n_0}, \cdots y_n\}$.

3.2.1 Filter initialization

Although the special-services segment of the total network is rapidly growing (at a rate of more than 9 percent per year) there are very few circuit groups or point-to-point disaggregated groups constantly growing. Most of them vary around a constant value, many of the groups have all their circuits disconnected (about 30 percent of the groups "die" over a 5-year period), and more new groups appear.

This frequent in-and-out activity rules out delaying the forecast until sufficient data is available to make accurate estimates of $\mathbf{X}_{n_0,n_0-1}$ and P_{n_0}. It is important to have initial state-vector estimates as soon as observations are available.

We considered three filter initialization methods for implementation in SSFS. Subsequent testing on actual data files (described in Section 3.3) was used to decide on the most appropriate one. Each method assumed that the length of every circuit group history is somewhere between 2 and 71 months.* As mentioned in Section 2.2, we look at quarterly average values of the demand for special services. The three methods are the following:

(*i*) Estimate $\hat{\mathbf{X}}_{n_0,n_0-1}$ and $\hat{P}_{n_0}$ by unweighted least squares. Assume a linear first-order model of the form $y = \beta_0 + \beta_1 z + \epsilon$, where z is the time variable (in our case, it is just the index of the observations, since the seasonal analysis can be assumed equally spaced in time), and ϵ, an error variable with mean zero and unknown variance σ^2. Given the observations $y = (y_0, y_1, \cdots, y_{n-1})$ taken at times $z = (0, 1, \cdots, n-1)$, y_n is estimated (by least squares) by $\hat{y}_n = \hat{\beta}_0 + \hat{\beta}_1 z_n$ and $\hat{y}_0 = \hat{\beta}_0$. Therefore,

$$\hat{y}_n = \hat{\beta}_0 + \hat{\beta}_1 z_n = \hat{\beta}_0 + \hat{\beta}_1 n = \hat{y}_{n-1} + \hat{\beta}_1 ,$$

$$\hat{x}^1_{n_0,n_0-1} = \hat{y}_n,\ \hat{x}^2_{n_0,n_0-1} = \hat{\beta}_1,\ \hat{p}^{11}_{n_0} = \text{var}\ \hat{y}_n ,$$

$$\hat{p}^{22}_{n_0} = \text{var}\ \hat{\beta}_1, \quad \text{and} \quad \hat{p}^{12}_{n_0} = \hat{p}^{21}_{n_0} = \text{cov}(\hat{y}_n, \hat{\beta}_1)$$

$$= \tfrac{1}{2}(\text{var}\ \hat{y}_n + \text{var}\ \hat{\beta}_1 - \text{var}\ \hat{y}_{n+1}).$$

The estimates $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{y}_n$, and $\hat{\sigma}^2$ are obtained with the usual regression formulas (as in Ref. 6).

(*ii*) Use the present SSFS model prediction for $\hat{\mathbf{X}}_{n_0,n_0-1}$ and estimate $\hat{P}_{n_0}$ from method (*i*).

(*iii*) Use the first quarter of data (2 or 3 months) to obtain a crude estimate $\hat{\mathbf{X}}_{1,1}$($\hat{x}^1_{1,1}$ = quarterly average, $\hat{x}^2_{1,1}$ = the slope of a line best fitting the data). Then use the Kalman filter algorithm sequentially on each quarterly average demand up to the current date. Estimate, as in

* When only one month is available, an arbitrary growth factor has to be used, and 71 months is the maximum history length stored by the SSFS history files.

(i), the initial error covariance matrix as the average of the errors in extrapolating for the 3 months in the second quarter.

Intuitively, this last method was expected to perform best, since, in most of the cases, enough history was available for the filter to achieve steady-state performance. Experimental testing of these initialization procedures indicated that indeed method (iii), the use of the filter algorithms as early in the history as possible, resulted in the best initial state-vector estimates $\hat{\mathbf{X}}_{n_0,n_{0-1}}$ and $\hat{P}_{n_0}$.

Weighted least squares, using minimum variance initialization estimates,[7] was not considered because special-services demand data may have many deterministic jumps. A distinction between these jumps and possible outliers could not be made since there were no records to indicate when such significant events occurred. This lack of information is equivalent to changing the characteristics of the vector $\mathbf{U}_n$ in eq. (5) into a random variable with unknown distribution.

3.2.2 Specification of R and Q

Various procedures exist for the estimation of the $\{R_n\}$ and $\{Q_n\}$ parameters. The methods vary in their relative complexity and the number of assumptions needed for the underlying statistical properties of the system. In most applications with relatively short time series, little improvement in performance is expected from a highly sophisticated specification procedure. A simpler method is used: instead of trying to identify $\{R_n\}$ and $\{Q_n\}$ for each individual time series, a scalar R and a matrix Q are determined that approximate the general nature of all series considered. Consequently, only one gain sequence $\{K_n\}$ and initialization matrix P_{n_0} has to be precomputed and applied to all circuit histories.

In our case, estimation of R and Q is obscured by the occurrence of deterministic events. For example, to estimate R, the series first has to be cleansed of special events, but any special events recognition is based on a measure of R. Nevertheless, upper bounds for the measurement error variance can be estimated assuming no deterministic events. The estimate measurement error, R, was found to be approximately 5 for Company B and 19 for Company C.

It can be shown[4,8] that the calculation of the gain sequence depends only on the relative magnitude of Q compared to R. Hence, if R is normalized to unity, only Q needs to be specified. We discuss the influence of R and Q on the gain sequence, filter responsiveness, and the selection of specific values for Q, based on experimental testing, in the next section.

3.2.3 The gain sequence

Accurate specification of the elements of Q is important, especially as they affect the gain sequence $\{K_n\}$ values for large n. Fig. 4 shows

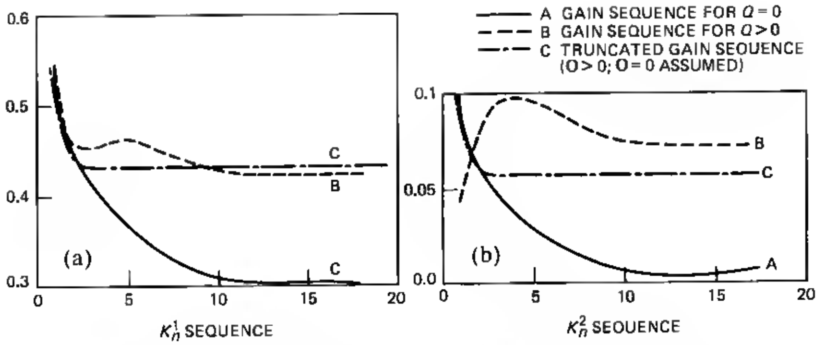

Fig. 4—Kalman filter gain sequences. (a) K_n^1 sequence; (b) K_n^2 sequence.

how the gain sequence is influenced by different assumptions made on the elements of Q. For $Q \equiv 0$ the gain sequence converges to zero, since $Q = 0$ is equivalent to a process $\{\mathbf{X}_n\}$ evolving deterministically relative to the initial set of parameters (see Fig. 4, curve A). A nonzero Q will force the sequence $\{K_n\}$ to give enough weight to new observations $\mathbf{y}_n$ so that a nonstationary process is correctly tracked by the filter (Fig. 4, curve B).

The choice of $\{K_n\}$ is based on the desire to be responsive to changes in demand, while maintaining relatively stable forecasts. To obtain this result even when the true statistical nature is not known and Q estimates are difficult to obtain, a truncated gain sequence can be used.[1,7] A truncated gain sequence K_n' (Fig. 4, curve C) is defined as

$$K_n' = \begin{cases} K_n & \text{if} \quad n \le n^* \\ K_{n^*} & \text{if} \quad n > n^* \end{cases} \qquad \text{and} \quad n^* \ge 1,$$

where K_n is calculated under the false assumption that $Q \equiv 0$, and n^* is empirically determined to ensure sufficient responsiveness and near-optimal filter performance. Another advantage in using the K_n' sequence over the optimal sequence is that a finite vector $\{K_1, \cdots, K_{n^*}\}$ can be computed and stored. [Fig. 4 is derived from eqs. (10), (11), and (12), and estimates of R, Q, and P_{n_0}.]

Given the demand data characteristics in our case, the matrix Q had to be nonzero to ensure filter responsiveness to random variation in the model parameters. For the normalized R_0 value of 1, different matrices Q were tested and corresponding steady-state gains calculated.

Figure 5 shows the theoretical performance of different gain sequences for the SSD-SPA algorithm when the true modeling error covariance matrix is constant and not zero ($Q \neq 0$): Curve A is the theoretical performance when the gain sequence is calculated under the false assumption that $Q_n \equiv 0$, curve B is obtained when the true Q

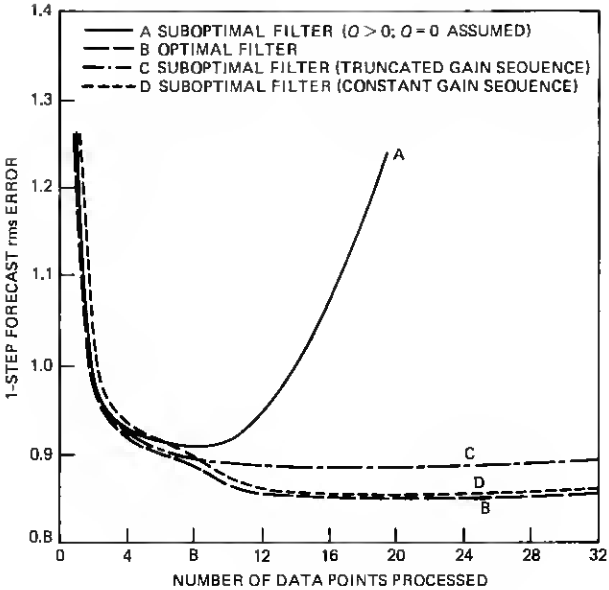

Fig. 5—Theoretical algorithm performance.

is used in obtaining the gain sequence, while curve C shows the theoretical performance of the truncated gain sequence ($n^* = 10$), also computed under the assumption that $Q_n \equiv 0$. This figure is derived from the generalized formula for S_n

$$S_n = (I - K_nH_n)P_n(I - K_nH_n)^T + K_nR_nK_n^T,$$

which is true for an arbitrary gain matrix K_n.[4,7]

The initial values of the gain sequence depend on the $\hat{P}_{n_0}$ values. The absolute values of the $\hat{P}_{n_0}$ were varied to obtain the transient gains (K_n, $n \leq 14$) that gave the best algorithm performance. This $\hat{P}_{n_0}$ and Q produced a near constant (over time) gain-vector sequence. Further testing on many time series indicated that a single algorithm using constant gains performs as well as any other. This empirical result is substantiated by theoretical near-optimal performance of a constant gain sequence, shown in Fig. 5, curve D. The constant gain was selected to be the best approximation obtained to the optimal gain. Consequently, constant gains were selected for the SSD-SPA implementation.

3.2.4 Outlier detection and data validation

An important characteristic of the special-service demand time series is the high volatility. This volatility affects the expected quality of the forecasting procedures and the determination of outlier detection screens and responses. Usually, outliers are defined as those measure-

ments that are significantly different from the trend because of data volatility, measurement errors, or deterministic changes in the level. Significance is determined based on tbe observed measurement variance about the assumed trend line (usually a band of 2 to 3 standard deviations around the expected line). Once a measurement falls beyond these boundaries, the outlier detection routine determines whetber the measurement indicates a change in the level of the trend or whether it is an outlier (data volatility or measurement error). The former case is decided based on subsequent measurements, i.e., if the following data conform to the cbange. In the latter case, the measurement has to be partially or complete ignored, based on the probability of being a true measurement of volatile demand or an erroneous one.

Previous studies indicated that the majority of the grouped demand time series is truly of a highly volatile nature with sudden changes; large and frequent changes (up to 1000 circuits added or subtracted in a single month), many times in opposite directions (big rise in level followed shortly by a big drop), were evident.

Consequently, it is impossible in the case of the special-services demand data to decide if an outlier was produced by a data-base error; therefore, no data validation decisions are recommended for the outlier detection routines.

Instead, such errors if present will be handled by the deterministic event detection and response procedure described in the next section.

3.2.5 Deterministic events detection and response

As mentioned in Section 2.1, two other important demand patterns must be considered in designing a projection algorithm:

(*i*) Significant changes in the demand level when subsequent observations confirm the supposition that a special event had taken place.

(*ii*) Zero growth when the time series remain for long periods of time at constant levels.

3.2.5.1 Step changes in demand levels. Since data is available monthly, detection of significant level changes should be made monthly even though the forecast is made quarterly. In this way, a quarterly response will be the result of at least three, and up to a maximum of six monthly movements. Let

$$d_i = y_{\text{month}_i} - y_{\text{month}_{i-1}},$$

$$\bar{d}_i = (y_{\text{month}_i} + y_{\text{month}_{i-1}})/2, \text{ and}$$

$d_i^* =$ maximum value $|d_i|$ can have before it is considered significant.

Two functional forms are usually used[1,3] to calculate d_i^*: the linear function $d_i^* = a + b\bar{d}_i$ or the mixed linear-exponential function $d_i^* =$

$\bar{d}_i(a + be^{c\bar{d}_i})$. In general, the latter form has the advantage that it increases the boundaries, percentagewise for small values of $\bar{d}_i$. For the special-services demand time series, the integral nature of the d_i's made this advantage insignificant. Optimum values for a, b, and c parameters were experimentally tested for both functional forms, but no improvement was found in the algorithm performance when the mixed linear-exponential boundaries were used. Since the simple linear form reduced the total computational time, we recommended the following linear deterministic event boundaries for SSD-SPA:

$$d_i^* = 0.7 + 0.11\ (y_{\text{month}_i} + y_{\text{month}_{i-1}}) \quad \text{for all} \quad i \geq 2.$$

We present next a brief description of the detection and response to past deterministic changes in the demand.

(*i*) *Detection step*

This procedure first determines if the given d_{i-1} is significant, and second if the subsequent d_i confirms this event. This confirmation means that either d_i has the same sign as d_{i-1}, or the net difference between $|d_i|$ and $|d_{i-1}|$ is large enough to be a deterministic event by itself. There are four possible cases: $d_{i-1} > 0$ and $d_i > 0$; $d_{i-1} < 0$ and $d_i < 0$; $d_{i-1} > 0$ and $d_i < 0$; and $d_{i-1} < 0$ and $d_i > 0$. In the first two cases, d_{i-1} is confirmed, since the next movement has the same sign. In the last two cases, the movements have opposite signs. To distinguish between volatility and special events, we subtract from both what can be attributed to volatility, i.e., minimum $\{|d_i|, |d_{i-1}|\}$. There are, then, two cases: $|d_{i-1}| > |d_i|$, and $|d_{i-1}| < |d_i|$.

Case 1: $|d_{i-1}| > |d_i|$. Then new $d'_{i-1} = d_{i-1} + d_i$ and new $d'_i = 0$. If d'_{i-1} compared to d^*_{i-1} is significant, then d'_{i-1} is a special event and $d'_i = 0$. If not, both d'_i and d'_{i-1} are zero.

Case 2: $|d_{i-1}| < |d_i|$. Then new $d'_i = d_i + d_{i-1}$ and new $d'_{i-1} = 0$. If $|d'_i| \geq d^*_i$, then d'_i is a special event and $d'_{i-1} = 0$. Otherwise both are zero.

(*ii*) *Response step*

From these monthly detected level changes, the quarterly events have to be calculated. A quarterly value Y_j is an average of three months: y_i, y_{i+1}, and y_{i+2}. (The Kalman filter model uses this Y_j as data, as described in paragraphs 3.2.1 and 3.2.2.) The effect of a monthly change on the quarterly average depends on the position of the month in that quarter. If the change happens in the first month of the quarter (d_i), then all y's in Y_j are moved to this second level, and the change in quarterly averages is $D_j = Y_j - Y_{j-1} = d_i$. If the change happens in the second month (d_{i+1}), then only y_{i+1} and y_{i+2} reflect the change, and $Y_j - Y_{j-1} = \frac{2}{3}d_{i+1}$. The remaining $\frac{1}{3}d_{i+1}$ will appear as a difference between Y_j and Y_{j+1}. Finally, if the change appears at $y_{i+2}, (d_{i+2})$, then $Y_j - Y_{j-1} = \frac{1}{3}d_{i+2}$ and $Y_{j+1} - Y_j = \frac{2}{3}d_{i+2}$.

Consequently, the changes observed on quarterly values are determined by five possible monthly events:

$$D_j = \tfrac{1}{3}d_{i-2} + \tfrac{2}{3}d_{i-1} + d_i + \tfrac{2}{3}d_{i+1} + \tfrac{1}{3}d_{i+2}\,.$$

Since we do not know how much of any D_j is actually normal growth, we recommend that when D_j fully explains the difference between Y_j and Y_{j-1}, to consider that it already included the growth. Certainly D_j is not available before Y_j, and therefore, the state estimate $\hat{x}^1_{j,j-1}$ is calculated first using the estimates $\hat{u}^1_{j,j-1}$ of future deterministic events which can be input to SSD-SPA, or from eq. 13:

$$\hat{\mathbf{X}}_{j,j-1} = \Phi\hat{\mathbf{X}}_{j-1,j-1} + \hat{\mathbf{U}}_{j,j-1}.$$

After D_j can be calculated ($D_j = \hat{u}^1_{j,j}$ = estimate of u^1_j after Y_j has been observed), then the state estimate $\hat{x}^1_{j,j-1}$ can be updated by:

$$\begin{aligned}
\hat{x}^1_{j,j-1} &\leftarrow \hat{x}^1_{j,j-1} - \hat{u}^1_{j,j-1} + D_j \quad \text{if} \quad D_j \neq Y_j - Y_{j-1} \quad \text{or} \quad D_j = 0 \\
&\leftarrow \hat{x}^1_{j,j-1} - \hat{u}^1_{j,j-1} + D_j - \hat{x}^2_{j-1,j-1} \quad \text{if} \quad D_j = Y_j - Y_{j-1} \quad \text{and} \quad D_j \neq 0.
\end{aligned}$$

Then, eq. 9 follows to calculate $\hat{\mathbf{x}}_{j,j}$.

Since events often do not occur as planned, this procedure also ensures algorithm recovery when erroneous estimates of future deterministic events are input to SSD-SPA.

3.2.5.2 Zero growth. Two quarters (or 6 months) with constant level of demand are regarded as sufficient evidence that the main tendency of that particular circuit group is to stay at that level for a longer period of time. However, if the filter estimate of the growth is not very close to zero, it may take many quarters to finally converge to zero since the filter has to be robust enough to perform on other higher volatile series. An appropriate procedure to force the growth estimate to converge to zero faster is to reduce the growth estimate ($\hat{x}^2_{n,n}$) by a factor γ (i.e., $x^2_{n,n}$ becomes $x^2_{n,n}/\gamma$) whenever zero quarterly growth is observed. Subsequent testing found $\gamma = 2$ to be a good value and concluded that this test is very robust for small variations of the γ parameter.

3.3 Performance analysis

Three objectives were identified for the SSD-SPA performance analysis: First, to determine and quantify the improvement in forecast accuracy, rms error, stability, and misplacements relative to the existing forecasting algorithm (described in Section 2.3). Second, to determine if the proposed algorithm has the desired properties (listed in Section 2.3.3) derived from the special-service demand data characteristics. Third, to assess the potential economic benefits resulting from incorporating SSD-SPA into SSFS versus its implementation costs.

The SSD-SPA was evaluated quantitatively using the accuracy, rms error, stability, and misplacement relative error statistics. To ensure relative algorithm performance analysis consistency, the same data was used as in the SSFS study (Sections 2.1 and 2.3.2). For these time series, equivalent consecutive forecasts were produced using the new sequential projection algorithm, and forecast performance measures were calculated.

Network aggregated error statistics were used in the selection of algorithm parameters, as well as in comparing the new SSD-SPA performance to the present algorithm.

The resulting forecasting algorithm was found to be robust over small variations of all parameters around their optimum values.

3.3.1 *Results: Accuracy, rms error, stability, misplacements*

Figure 6 displays graphically the performance of the SSD-SPA using both companies' history data. Special-services demand sequential projection algorithm generates forecasts that are significantly more accurate and stable. Tables I and II give the SSD-SPA versus present algorithm relative improvements in forecast accuracy, rms error, stability, and total misplacement.

Figures 7a and 7b present two examples of the SSD-SPA versus present algorithm total error (TE) and misplacement (M) relative improvements for the forecasts generated in 1974.

Figure 8 gives histograms of relative improvement for the 1-year-ahead forecast accuracy, rms error, total misplacements, and stability

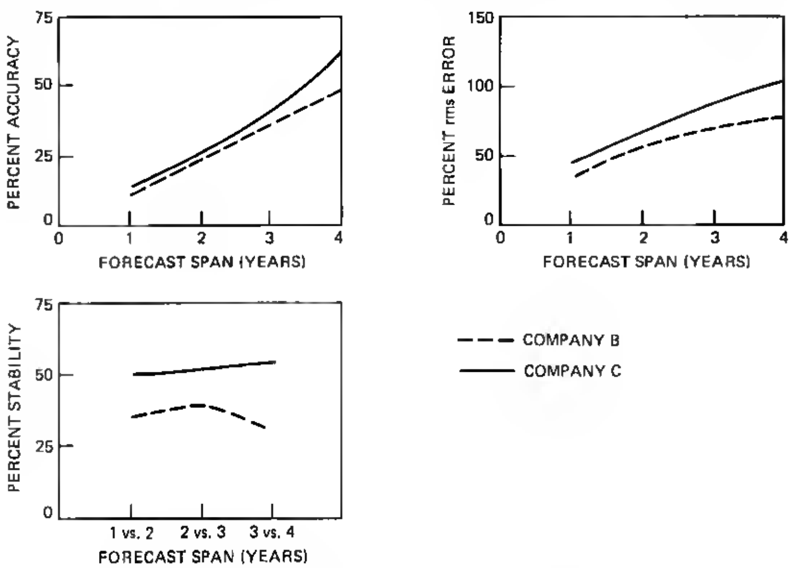

Fig. 6—SSD-SPA network average forecasting performance.

Table I—SSD-SPA Percent relative forecast improvement over present algorithm

Forecast Attribute	Company	1-Year Span	2-Year Span	3-Year Span	4-Year Span
Accuracy	C	30.0	29.6	29.6	31.9
	B	25.0	24.0	25.0	29.0
Root-mean-square	C	17.8	18.3	20.7	23.0
	B	19.7	19.1	22.2	24.0
Stability	C	15.0	27.0	38.8	
	B	20.0	33.0	51.0	

Table II—SSD-SPA Percent relative improvement in total misplacements over present algorithm

Company	Base Year Forecast	Forecasted Year 1975	1976	1977	1978
C	1974	17.4	19.5	20.2	21.6
	1975		18.3	15.7	13.7
	1976			18.3	13.1
B	1974	21.9	19.2	19.0	23.3
	1975		17.8	19.0	22.4
	1976			22.9	22.3

(1 versus 2-years-ahead for stability). Much of the observed improved performance is because the new algorithm can detect and properly respond to step changes in the demand level. Figs. 9a and 9b show how SSD-SPA processes the data shown previously in Figs. 3a and 3b as examples of SSFS poor performance.

It should be noted that both examples only show how past deterministic events (before the start of the forecasting period, i.e., before July, 1974, in Fig. 9a, and July, 1975, in Fig. 9b) are treated. No knowledge was assumed about future special events, such as the one on October, 1976 (Fig. 9b). Once the data up to these events are available, even if no, incomplete, or wrong information would be input into SSD-SPA, the algorithm could recognize them and properly adjust the forecast, as was shown for the events on May, 1974 (Fig. 9a) and February, 1974 (Fig. 9b). The present SSFS algorithms treated these events as part of the normal growth, as shown in Figs. 3a and 3b.

3.3.2 Small integer forecast

In Section 2.3.3, we stated six desirable properties for the new forecasting algorithm based on the demand time series characteristics.

The unequal weigbting of data, acceptance of exogenous information, and a short initialization period are shown to be part of the proposed mathematical model itself (Section 3.1). The recursive filter model adds computational efficiency since it does not explicitly use past data. Recognition of the past deterministic events and algorithm

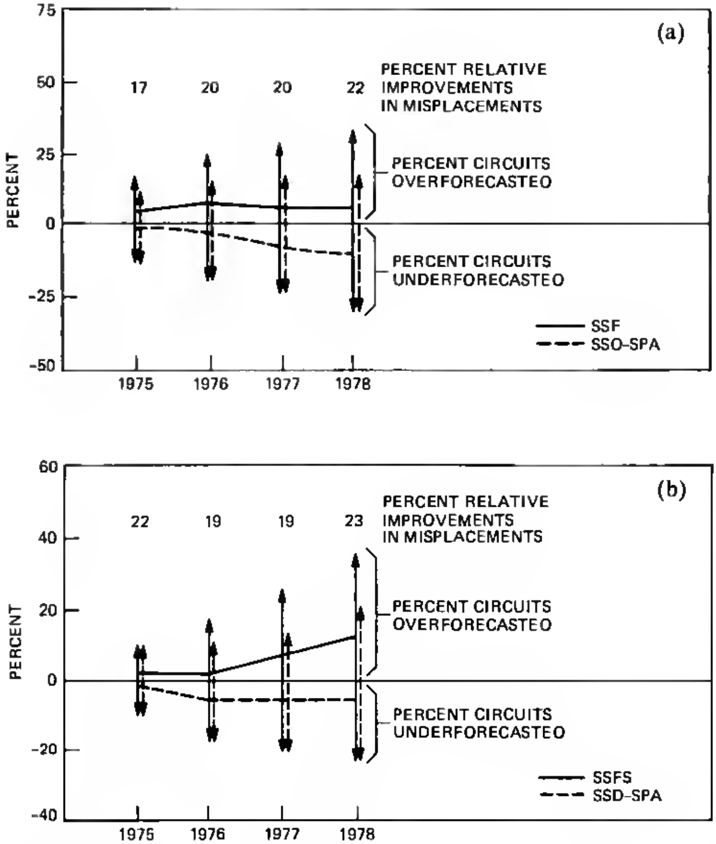

Fig. 7—SSD-SPA versus SSFS: total error and misplacement (forecasts generated in 1974). (a) Company C; (b) Company B.

recovery are ensured by the two procedures described in Section 3.2.5. It only remains to see if SSD-SPA performs adequately when small integers are to be forecasted.

To quantify this, the tests were repeated using only those point-to-point time series consisting of integers less than 10 (approximately 80 percent of all point-to-point demand time series).

Results of these tests on both companies' data bases showed that for small integers the relative forecast improvement of SSD-SPA is about 50 percent in accuracy, 30 percent in rms error, 30 to 66 percent in stability, and 50 percent in total misplacement. Moreover, total forecast error was found to range between 1 to 3 percent for SSD-SPA versus 3 to 28 percent for the present method. These last results excluded the "vanishing" time series in order to obtain unbiased attribute estimates.

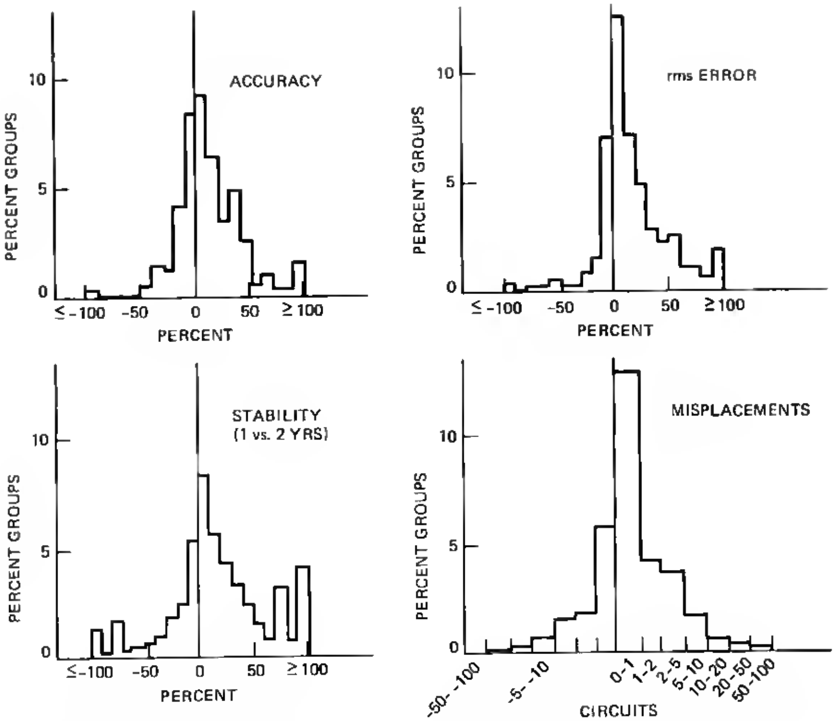

Fig. 8—SSD-SPA versus SSFS: percent relative improvements in 1-year forecast attributes for Company C.

3.3.3 Economic benefits and implementation costs

The comparative study of the present SSFS forecasting algorithm and the SSD-SPA showed that the new algorithm generates forecasts that are significantly more accurate and stable. Implementation of SSD-SPA in SSFS would, therefore, translate into important economic benefits in three areas: capital expenditures, forecasters' time, and electronic data processing costs.

(*i*) The major impact is expected to be on capital savings. The following analysis is based on the SSFS preliminary forecast before any manual adjustments are made. (There are no records available with the final adjusted forecasts made at different times in the past, nor with the exogenous information available to the forecaster.) The results showed the 1-year SSFS forecast positive misplacement of circuits to be 12 percent, on the average. That is, 12 percent of the total special-services circuits in the 1-year forecast could be in the wrong groups resulting in an overprovisioning. One-year results are used to be conservative; additionally, for a 1-year error there is less chance to reuse misplaced facilities.

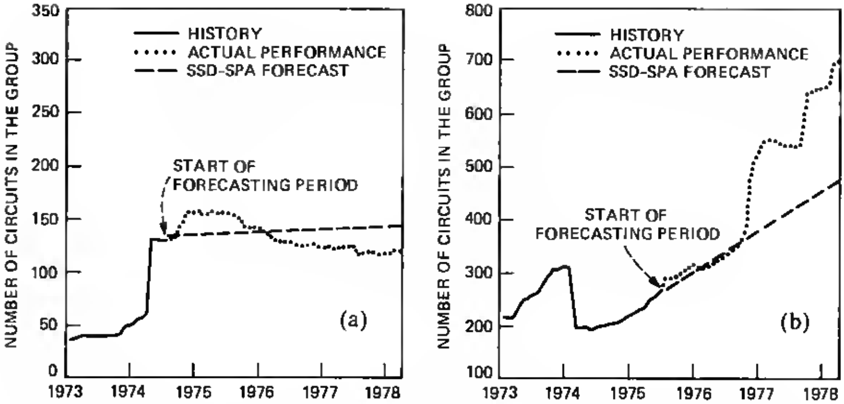

Fig. 9—Circuit groups with deterministic jumps. SSD-SPA treatment. (a) Example 1; (b) Example 2.

The SSD-SPA reduces the misplacement to 7 percent; a reduction of 5 percent. Theoretically, 5 percent of the total special-services circuit network could be removed without a change in service. Underprovisioning is approximately the same for both algorithms.*

(*ii*) The improved forecast accuracy, the recognition of past deterministic events, and the shorter forecast initialization requirements are SSD-SPA features that translate into fewer manual forecast adjustments. Fewer adjustments would permit the forecasters to concentrate more of their efforts to follow the economic conditions and estimate their impact on the future demand for special services.

(*iii*) The SSD-SPA is based on one forecasting model only and makes no explicit use of all the data history. Consequently, run times and core usage would be reduced. Although the absolute savings are not large, they would make SSFS very suitable for an on-line use.

IV. CONCLUSIONS

The goal of our work was to design an algorithm able to forecast future demands for special services: highly volatile time series mainly consisting of small integers, and with numerous deterministic jumps. We have shown that a linear, dynamic time-series model with linear growth and deterministic input, together with the Kalman filtering technique for state vector estimation and prediction, can produce demand forecasts which are significantly more accurate and stable

* This apparent positive bias is due to two types of groups. The first is those groups which "vanish" during the period. The second is those in which large deterministic events occurred. In the new algorithm, these events can be handled by input of marketing information.

than the forecasts produced by the best (highest R^2) choice of four unweighted regression models: the linear, exponential, and first- and second-order autoregressive. The new model, its attributes, and specific parameters were selected based on the characteristics of actual special-service demand history from three BOCS.

The improvement in accuracy is due to the capability of the system to track nonstationary processes, and also to recognize and react properly to deterministic changes in the demand, even when no, or wrong exogenous input was available. The use of a single model is responsible for much of the stability improvement. Additionally, SSD-SPA can produce many views of future demand using different assumptions on future events, it requires a short initialization period, and it results in the need for fewer manual adjustments. Therefore, we propose to replace the existing algorithm in the SSFS by this simple and more efficient algorithm.

V. ACKNOWLEDGMENTS

The author wishes to thank J. P. Moreland, S. R. Neal, and B. A. Whitaker for their helpful comments and suggestions. Special thanks go to R. Troester for the time and affort spent to modify SSFS, and to D. E. Sohn for help in obtaining and preparing the data bases. The curves of Fig. 5 were obtained from a modified version of C. R. Szelag's program.

REFERENCES

1. A. J. David and C. D. Pack, "The Sequential Projection Algorithm: A New and Improved Traffic Forecasting Procedure," Proc. Ninth Int'l. Teletraffic Congress, Torremolinos, Spain, 1979.
2. C. D. Pack and B. A. Whitaker, "Kalman Filter Models for Network Forecasting," B.S.T.J., this issue.
3. J. P. Moreland, "A Robust Sequential Projection Algorithm for Traffic Load Forecasting," B.S.T.J., this issue.
4. A. Gelb, ed., *Applied Optimal Estimation*, MIT Press, Cambridge, Mass., 1974.
5. A. M. Jazwinski, *Stochastic Processes and Filtering Theory*. Academic Press, New York, 1970.
6. N. R. Draper and H. Smith, *Applied Regression Analysis*, John Wiley and Sons, N.Y., 1966.
7. C. R. Szelag, "A Short-Term Forecasting Algorithm for Trunk Demand Servicing," B.S.T.J., this issue.
8. R. E. Kalman, "A New Approach to Linear Filtering and Prediction Problems," J. Basic Eng., *82*, pp. 34–45, March, 1960.